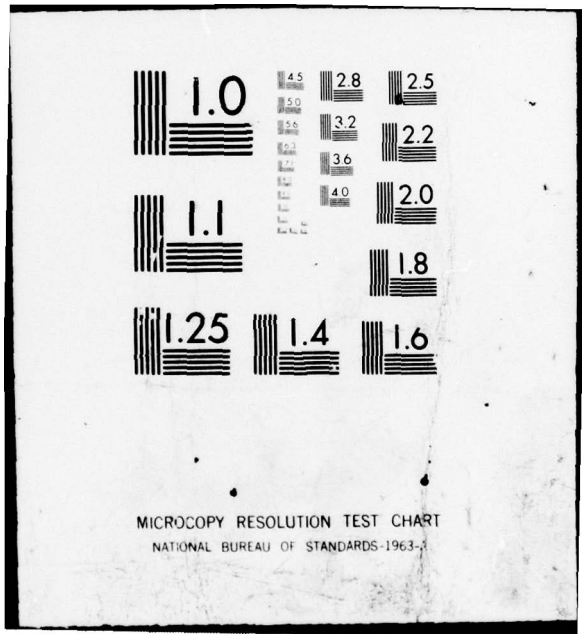
AD-A079 735 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
ON SOLVING ROBUST AND GENERALIZED LINEAR REGRESSION PROBLEMS.(U)
SEP 79 D M GAY DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-2000 NL

END
DATE FILMED
2-80
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MRC Technical Summary Report # 2000

(3)

ON SOLVING ROBUST AND GENERALIZED LINEAR REGRESSION PROBLEMS

David M. Gay

LEVEL II

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

September 1979

(Received July 16, 1979)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, D. C. 20550

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ON SOLVING ROBUST AND GENERALIZED LINEAR REGRESSION PROBLEMS*

David M. Gay

Technical Summary Report #2000
September 1979

ABSTRACT

A number of regression problems require finding a parameter vector x^* that minimizes an objective function of the form $\sum_i \rho_i(r(x))$, where $r_i(x)$ is the ith component of the (generalized) residual vector $r(x)$ associated with the problem and ρ_i is the ith criterion function. The examples given include (linear and nonlinear) least squares, robust regression, logistic regression, and Poisson regression. These problems have a common structure which may often be worth exploiting, especially in cases where $r(x)$ is a nonlinear function of x. We study this structure and how to use it. This provides an opportunity to discuss some ideas applicable to general unconstrained optimization and, in particular, to point out the advantages of a model/trust-region approach. For nonlinear $r(x)$, we recommend generalizations of some techniques that have proven worthwhile on nonlinear least-squares problems in which the optimal residual vector $r(x^*)$ may be either large or small. Appendices consider the numerical linear algebra involved in computing a trial change to the parameter vector x and give a new perturbation theorem on linear least squares that is useful in analyzing the numerical behavior of the procedure recommended for computing this trial change.

AMS(MOS) Subject Classifications: 15-04, 15A09, 65F05, 65F25, 65K05, 90C30

Key Words: Unconstrained optimization, regression, numerical linear algebra, roundoff analysis

Work Unit Number 5 - Mathematical Programming and Operations Research

---

*Presented at "Nonlinear Optimization and Applications," L'Aquila, Italy, 18-20 June 1979.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024. This material is based upon work supported by the National Science Foundation under Grant No. MCS78-09525.

SIGNIFICANCE AND EXPLANATION

Many researchers employ mathematical models. Most models contain parameters, which may be chosen to make the model fit the available data as well as possible (in a sense that depends on the model). In this paper we consider the problem of choosing the parameters for a common class of models in which the desired parameter vector x^* minimizes an (unconstrained) objective function, of the form $\sum_i \rho_i(r_i(x))$, where r_i is the ith (generalized) residual of the model and ρ_i is a scalar criterion function. (Often r_i is the model's error at the ith observation.) We briefly give some examples of such problems, then discuss ways to exploit the common structure that these problems share. This leads us to discussing strategies for solving general unconstrained minimization problems and to point out the advantages of using a so-called "model/trust-region approach," wherein the change made in the current parameter estimate is chosen so as to approximately minimize a local model of the objective function on an estimate of the region about the current iterate where this local model is reliable. For problems in which the residual vector $r(x)$ is a nonlinear function of x, we recommend generalizations of some techniques that have proven worthwhile in nonlinear least-squares problems in which the optimal residual vector $r(x^*)$ may be either large or small.

ON SOLVING ROBUST AND GENERALIZED LINEAR REGRESSION PROBLEMS*

David M. Gay

1. Introduction

A number of regression problems require finding a parameter vector x^* which minimizes an objective function $\varphi : \mathbb{R}^p \to \mathbb{R}$ of the form

$$\varphi(x) = \sum_{i=1}^{n} \rho_i(r_i(x)) \quad , \tag{1.1}$$

where $r_i : \mathbb{R}^p \to \mathbb{R}$, $\rho_i : \mathbb{R} \to \mathbb{R}$, and $n \geq p$. Often $r(x) := (r_1(x), \cdots, r_n(x))^T$ is the residual vector corresponding to a linear or nonlinear model; we shall therefore call it the (generalized) "residual vector" even in cases where minimizing φ is not intended to make $r(x^*)$ near zero. In the case of robust regression, ρ_i is sometimes called a robust loss function. But "loss" has other meanings, so we shall refer to ρ_i as the ith criterion function.

Problems of the form (1.1) have in common a structure which, we suspect, it will often prove worthwhile to exploit, especially when the residual vector $r(x)$ is a nonlinear function of x. After giving examples of (1.1) in the next section, we restrict our attention in §3 to the common case where $r(x)$ is an affine function of x. This gives us the opportunity to discuss some ideas applicable to general unconstrained optimization and, in particular, to point out the advantages of a model/trust-region approach. In §4 we turn to the general case where $r(x)$ is nonlinear and recommend trying the obvious generalizations of some ideas that have proven worthwhile for nonlinear least squares. Appendix A deals with the numerical linear algebra of solving the special linear least-squares problems that arise when computing certain "Newton" steps, and Appendix B states and proves a perturbation theorem used in Appendix A.

---

*Presented at "Nonlinear Optimization and Applications," L'Aquila, Italy, 18-20 June 1979.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024. This material is based upon work supported by the National Science Foundation under Grant No. MCS78-09525.

2. Examples

Perhaps the most common problem of the form (1.1) is the least-squares problem, in which $\rho_i(\tau) = \frac{1}{2}\tau^2$ for all i. This problem may arise, for example, when one has a model $f_i(x)$ that predicts what experimental response will be obtained under the ith set of conditions, and one measures response y_i under these conditions; if there are independent, normally distributed errors in the y_i, then the maximum-likelihood estimate x^* for the model parameters minimizes (1.1) with $r_i(x) = f_i(x) - y_i$ and $\rho_i(\tau) = \frac{1}{2}\tau^2$.

The least-squares parameter estimate can be strongly influenced by errors in the data y_i, so a number of so-called "robust regression" techniques have been proposed for obtaining good estimates of the model parameters in cases where some of the y_i contain large errors. One such approach involves solving (1.1) with ρ_i a robust criterion function and $r_i(x) = (f_i(x) - y_i)/\sigma$, where σ is a scale parameter that has been determined by some other means. A number of robust criterion functions have been proposed, such as the eight choices considered by Holland and Welsch in [HolW77]. These include the criterion functions of Huber [Hub64],

$$\rho_i(\tau) = \begin{cases} \frac{1}{2}\tau^2 & \text{if } |\tau| \le H \\ (|\tau| - H/2)H & \text{if } |\tau| > H \end{cases} , \tag{2.1}$$

of Fair [Fai74],

$$\rho_i(\tau) = F[|\tau| - F \cdot \ln(1 + |\tau|/F)] , \tag{2.2}$$

of Welsch [DenW78],

$$\rho_i(\tau) = \frac{1}{2}W^2[1 - \exp(-\tau^2/W^2)] , \tag{2.3}$$

and of Hinich and Talwar [HinT75],

$$\rho_i(\tau) = \begin{cases} \frac{1}{2}\tau^2 & \text{if } |\tau| \le T \\ T^2/2 & \text{if } |\tau| > T \end{cases} . \tag{2.4}$$

On problems with small residuals, all of these cirterion functions behave very much like the least-squares criterion function. The tuning constants H, F, W, T of (2.1-4) may be chosen to make (1.1) yeild parameter estimates with a specified level of efficiency - see [HolW77].

The generalized linear models of Nelder and Wedderburn [NelW72] give rise to other problems of the form (1.1). In addition to the least-squares problem, these include logistic and Poisson regression.

The logistic regression problem arises if one models the probability π_i of success in an experiment under the ith set of conditions by $\pi_i/(1 - \pi_i) = e^{r_i(x)}$. If the experiment is repeated ν_i times under these conditions and σ_i successes occur, then (under reasonable assumptions) the probability of the observed outcomes is $\prod_{i=1}^{n} \pi_i^{\sigma_i}(1 - \pi_i)^{\nu_i - \sigma_i}$, and choosing x^* to maximize (the logarithm of) this probability amounts to minimizing (1.1) with

$$\rho_i(\tau) = \nu_i \ln(1 + e^{\tau}) - \tau \sigma_i \quad . \tag{2.5}$$

Analogous reasoning motivates the Poisson regression problem. Suppose one observes n independent Poisson processes and obtains a count of ν_i for the ith process, an event of probability $e^{-\lambda_i}\lambda_i^{\nu_i}/(\nu_i!)$. If one models $\lambda_i = e^{r_i(x)}$, then finding a maximum likelihood estimate x^* for x amounts to minimizing (1.1) with

$$\rho_i(\tau) = e^{\tau} - \nu_i \tau \quad . \tag{2.6}$$

With the exception of (2.4), all of the above criterion functions are continuously differentiable. The discontinuity in the ρ_i' of (2.4) causes no serious trouble for the iterative schemes considered below, at least so long as $r(x)$ is continuously differentiable, since (as is easily seen), points of discontinuity are not points of attraction for these schemes. The discontinuity in the Huber ρ_i'' (2.1) causes no problems either, so we will refrain from further discussion of the continuity of the criterion functions and will assume them to be at least twice differentiable in what follows.

3. Linear Problems

Often when the problems sketched in §2 arise, $r(x)$ is a linear or affine mapping:

$$r(x) = Ax - b \quad , \tag{3.1}$$

where $A \in \mathbb{R}^{n \times p}$ (i.e., A is an $n \times p$ matrix) and $b \in \mathbb{R}^n$. In this case, the gradient and Hessian of φ have particularly simple forms:

$$\nabla\varphi(x) = A^T \rho'(r(x)) \quad \text{and} \tag{3.2}$$

$$\nabla^2 \varphi(x) = A^T D(\rho''(r(x)))A \quad , \tag{3.3}$$

where $\rho'(r(x)) = [\rho_1'(r_1(x)), \cdots, \rho_n'(r_n(x))]^T$ and $D(\rho''(r(x))) = \operatorname{diag}(\rho_1''(r_1(x)), \cdots, \rho_n''(r_n(x)))^T$ is the diagonal matrix whose ith diagonal element is $\rho_i''(r_i(x))$.

Since (3.3) has such a simple form, it is reasonable to consider using the damped Newton's method

$$x^{k+1} = x^k - \lambda_k \nabla^2 \varphi(x^k)^{-1} \nabla\varphi(x^k) \tag{3.4}$$

to construct a sequence of iterates which, under reasonable conditions, converge to a (possibly local) minimizer x^* of (1.1). In (3.4), λ_k is a step length parameter chosen to assure $\varphi(x^{k+1}) < \varphi(x^k)$ when $\nabla\varphi(x^k) \neq 0$. It is unnecessary to choose λ_k so as to (nearly) minimize $\varphi(x^{k+1})$; indeed, it is nowadays generally recognized as inefficient to attempt this. But it is important to make $\varphi(x^k) - \varphi(x^{k+1})$ large enough that the iterates do not converge to a noncritical point of φ. See [Pow71], [GilM74], and §6.3 of [DenS79] for discussion of efficient ways to do this.

If ρ_i is the least-squares criterion function, A has rank p, and $\lambda_1 = 1$, then (3.4) converges in one iteration (when exact arithmetic is used). In this case, (3.4) amounts to the normal equations: $x^* = (A^T A)^{-1} A^T b$. When finite-precision arithmetic is used, explicitly computing $A^T A$ and $A^T b$ and solving $A^T A x^* = A^T b$ works well if A is well conditioned, i.e., if the condition number $\kappa = (\|A^T A\| \, \|(A^T A)^{-1}\|)^{1/2}$ of A is not too large. For large values of κ it is

usually much more accurate to employ a QR factorization of A, i.e., to factor A as QR, where $Q \in \mathbb{R}^{n \times p}$ has orthonormal columns and $R \in \mathbb{R}^{p \times p}$ is upper triangular, and then solve $R\,x^* = Q^T b$ - see [LawH74]. (This costs roughly $n\,p^2 - p^3/3$ multiplications and a similar number of additions, versus roughly $n\,p^2/2 + p^3/6$ for explicitly solving the normal equations, so using the QR factorization less than doubles the arithmetic overhead).

For other criterion functions having $\rho_i'' > 0$ for all i, such as (2.2), (2.5), and (2.6), we can readily convert the task of computing the Newton direction $-\nabla^2 \varphi(x^k)^{-1} \nabla\varphi(x^k)$ into that of solving a linear least-squares problem by setting

$$A^k = \mathcal{D}(\rho''(r(x^k))^{1/2})A \quad \text{and} \tag{3.5a}$$

$$b^k = \mathcal{D}(\rho''(r(x^k))^{-1/2})\rho'(r(x^k)) \quad , \tag{3.5b}$$

where $\mathcal{D}(\rho''(r)^{\pm 1/2}) = \mathrm{diag}(\rho_1''(r_1)^{\pm 1/2}, \cdots, \rho_n''(r_n)^{\pm 1/2})$, so that $\nabla^2 \varphi(x^k) = (A^k)^T A^k$ and $\nabla\varphi(x^k) = (A^k)^T b^k$. Using a properly computed QR factorization of A^k will then usually lead to a more accurately computed value for the Newton direction.

There are three commonly used ways to compute a QR factorization for use in solving linear least-squares problems: triangularization via Householder transformations (elementary reflectors), triangularization via (standard or fast) Givens transformations (plane rotations), and the stabilized Gram-Schmidt process - see [LawH74]. While all three yield similar numerical accuracy on a broad range of problems, use of the first two techniques on (3.5) can sometimes lead to a severe loss of accuracy in cases when $\rho''(r)$ has a component near zero. This point is discussed in detail in Appendix A, where it is shown that the stabilized Gram-Schmidt process does not share this drawback.

Some cost functions, such as (2.3), can have $\rho_i''(r_i(x)) < 0$ for certain i, so it is possible for the Newton direction $-\nabla^2 \varphi(x)\nabla\varphi(x)$ to be uphill, nonexistent, or nearly orthogonal to the gradient $\nabla\varphi(x)$. One way to overcome this difficulty is to

replace $\nabla^2 \varphi(x^k)$ in (3.4) by a positive definite matrix H^k, so that (3.4) becomes

$$x^{k+1} = x^k - \lambda_k (H^k)^{-1} \nabla\varphi(x^k) \quad . \tag{3.6}$$

For solving robust regression problems, Huber [Hub75] has considered choosing

$$H^k = A^T A \quad . \tag{3.7}$$

Since H^k remains the same for all k, this choice lets one avoid computing new matrix factorizations after the first iteration. But it often converges slowly. Holland and Welsch [HolW77] advocate using an H^k that is generally attributed to Beaton and Tukey [BeaT74]:

$$H^k = A^T \mathcal{D}(w(r(x^k)))A \quad , \text{ where} \tag{3.8a}$$

$$w(r) = (\rho_1'(r_1)/r_1, \cdots, \rho_n'(r_n)/r_n)^T \quad . \tag{3.8b}$$

This converges more rapidly than (3.7) but more slowly than Newton's method (near the solution) [HolW77]. (Since $\rho_i'(0) = 0$ for robust criterion functions, (3.8) amounts to replacing $\rho_i''(r_i)$ in (3.3) by the finite difference $w_i(r_i) = \dfrac{\rho_i'(r_i) - \rho_i'(0)}{r_i - 0} > 0$.) With this choice of H^k, it is again possible to compute the step direction, $-(H^k)^{-1} \nabla\varphi(x^k)$, by linear least squares, since $\nabla\varphi(x^k) = A^T \mathcal{D}(w(r(x^k)))r(x^k)$; when done with $\lambda_k = 1$, this is known as <u>iteratively reweighted least-squares</u>. Byrd and Pyne [ByrP79] have proven the remarkable result that $\lim_{k\to\infty} \nabla\varphi(x^k) = 0$ for <u>any</u> x^0 with this approach. Because robust regression problems are likely to have a number of local minima, Holland and Welsch [HolW77] recommend starting from an x^0 that minimizes $\|Ax - b\|_1 = \sum_{i=1}^{n} |(Ax - b)_i|$.

In connection with general unconstrained minimization, a number of other choices for H^k have been considered in the literature. Greenstadt, for instance, has proposed replacing negative eigenvalues of $\nabla^2 \varphi(x^k)$ by their absolute values [Gre67]. Murray has proposed an interesting modified Cholesky factorization [Mur72]. And Moré and

Sorensen [MorS77] have proposed modifying the eigenvalues of the block diagonal matrix in the Bunch-Parlett [BunP71] factorization of $\nabla^2 \varphi(x^k)$. Unfortunately, all of these schemes can generate search directions $-(H^k)^{-1} \nabla\varphi(x^k)$ which are nearly orthogonal to the gradient in certain cases when $\nabla^2 \varphi(x^k)$ is poorly conditioned. This can cause slow convergence.

Secant update methods (see [DenM77], where they are called quasi-Newton methods) provide another way to generate a reasonable positive definite substitute H^k for $\nabla^2 \varphi(x^k)$. They offer the advantage of less arithmetic overhead per iteration than methods which deal explicitly with $\nabla^2 \varphi$, but they also converge more slowly than these methods. They should be faster than (3.7), and at this writing it is unclear how they fare in comparison with (3.8).

Starting with Fletcher and Freeman [FleF75] and McCormick [McC77], there has been considerable interest of late in exploiting negative curvature. The idea is to replace the single search direction $-(H^k)^{-1} \nabla\varphi(x^k)$ in (3.6) by what Sorensen [Sor77] has termed a descent pair (s^k, d^k), in which $s^k = -(H^k)^{-1} \nabla\varphi(x^k)$ for some positive definite H^k and $(d^k)^T \nabla^2 \varphi(x^k) d^k \leq 0$, with equality only if $\nabla^2 \varphi(x^k)$ is positive semidefinite. In such a scheme, $x^{k+1} - x^k$ is a linear combination of s^k and d^k: Goldfarb [Gol77] suggests a combination of the form $\alpha^2 s^k + \alpha d^k$, while Moré and Sorensen [MorS77] prefer $\alpha s^k + \alpha^2 d^k$. The choice of s^k recommended in [Gol77] and [MorS77] can be nearly orthogonal to the gradient in some cases (and the higher the arithmetic precision, the more nearly orthogonal s^k and the gradient can be), so these schemes need further study.

The model/trust-region approach offers a more elegant way of dealing with cases where $\nabla^2 \varphi$ may not be sufficiently positive definite. This approach can generate steps in a direction of negative curvature (i.e., $(x^{k+1} - x^k)^T \nabla^2 \varphi(x^k)(x^{k+1} - x^k) < 0$ is possible), but in some ways it is simpler than the schemes that deal explicitly with negative curvature. Moreover, it avoids the slow convergence that can result from search directions nearly orthogonal to the gradient, and it can converge faster

than (3.8), because it reduces naturally to Newton's method near a strong local minimizer (one where $\nabla^2 \varphi$ is positive definite).

Among the first to consider the model/trust-region approach were Marquardt [Mar63], who did so in connection with nonlinear least-squares, and Goldfeld, Quandt, and Trotter [GolQT66], who considered general unconstrained optimization problems. (Levenberg, whose work [Lev44] is often mentioned in the same breath with Marquardt's, actually considered a different approach.) These authors did not attempt to control the trust-region radius δ^k directly, but worked instead with the Lagrange multiplier λ_*^k mentioned below. By contrast, Powell specified δ^k explicitly in [Pow70a-d]. When H^k is positive definite, his dogleg strategy produces an easily computed approximation to the step discussed below, but this strategy is of no help when $H^k = \nabla^2 \varphi(x^k)$ and $\nabla^2 \varphi(x^k)^{-1} \nabla \varphi(x^k)$ is a direction of negative curvature or is undefined. Hebden [Heb73] appears to have been the first to give a practical algorithm (and computer program) using the model/trust-region approach in the form which will now be described.

This approach works as follows. We pick a convenient model $q^k(s)$ for $\varphi(x^k + s)$ and choose a neighborhood T^k, called the trust-region, in which we believe q^k to be reliable. We then compute a candidate step s^k by (approximately) minimizing $q^k(s)$ on T^k (or on an approximation to T^k). If $\varphi(x^k + s^k) < \varphi(x^k)$, then $x^{k+1} = x^k + s^k$; otherwise $x^k = s^k$ and we choose a smaller trust-region T^{k+1}. This descent rule (together with a reasonable choice for T^{k+1}) makes convergence to a local minimizer quite likely.

In this section we consider only one choice for q^k, a quadratic model of the form

$$q^k(s) := \varphi(x^k) + s^T \nabla \varphi(x^k) + \frac{1}{2} s^T H^k s^k \quad , \tag{3.9}$$

in which H^k is either $\nabla^2 \varphi(x^k)$ or some approximation to it. The choices for T^k usually considered depend on a step bound δ^k (which will be discussed later) and

have the form

$$T^k = \{s \in \mathbb{R}^p : \|Ds\| \leq \delta^k\} \quad , \tag{3.10}$$

where D is a diagonal matrix having $D_{ii} > 0$. While choosing $D = I$, the identity matrix, works well on well-scaled problems, a proper choice of the scale matrix D can lead to significantly faster convergence on problems where the components of x are expressed in sharply contrasting units. For the linear problems of present concern, choosing D_{ii} to be some convenient norm of the ith column of A is usually reasonable.

If the norm $\|\cdot\|$ in (3.10) is the max-norm (i.e., $\|y\| = \|y\|_\infty = \max\{|y_i| : 1 \leq i \leq p\}$, which at first is appealing if one wishes to consider generalizing to problems having simple bounds, such as nonnegativity constraints, on some components of x), then computing the optimal s^k amounts to solving a quadratic programming problem with particularly simple constraints. Unfortunately, if H^k has some negative eigenvalues, then this quadratic programming problem can have a number of local minima (as many as 2^p).

Even when H^k is positive definite, it is often possible to compute s^k more rapidly or more accurately if $\|\cdot\|$ in (3.10) is the 2-norm (i.e., $\|y\| = \|y\|_2 = (y^T y)^{1/2}$). In this case, any s^k that minimizes $q^k(s)$ subject to $s \in T^k$ satisfies

$$[H^k + \lambda_*^k D^2]s^k = -\nabla\varphi(x^k) \quad , \tag{3.11}$$

where $\lambda_*^k \geq 0$, λ_*^k makes $H^k + \lambda_*^k D^2$ positive semidefinite, and $\lambda_*^k = 0$ if $\|D\, s^k\|_2 < \delta^k$ (see [Gay79]). In practice, if λ_*^k is positive, then it usually cannot be determined exactly, and attempting to approximate it with high accuracy would be inefficient. Good performance is usually obtained in this case if s^k is chosen to satisfy (3.11) with λ_*^k replaced by any λ^k that yields $\alpha\,\delta^k \leq \|D\, s^k\|_2 \leq \beta\,\delta^k$ for some fixed $\alpha < 1$ and $\beta > 1$. (Hebden [Heb73] and Moré [Mor78] choose $\alpha = 0.9$ and $\beta = 1.1$, while Dennis and Schnabel [DenS79] suggest $\alpha = 0.75$ and $\beta = 1.5$. In the

context of NL2SOL [DenGW79], we felt that the former choice gave slightly better performance than the latter.) An acceptable λ^k can usually be obtained in one or two tries using an iteration proposed independently by Hebden [Heb73] and Reinsch [Rei71] together with Moré's variation [Mor78] of Hebden's safeguarding scheme. A simple way of detecting and handling the exceptional case where $H^k + \lambda_*^k D^2$ is nearly (or actually) singular is described in [Gay79].

In cases where $\rho_i''(r_i) > 0$ for all i and $H^k = \nabla^2 \varphi(x^k)$, it is possible to avoid computing $\nabla^2 \varphi(x^k)$ explicitly and to work with a QR factorization of $J(\rho''(r(x^k)))^{1/2} A$ by carrying out calculations in the way described by Moré [Mor78]. The discussion above and in Appendix A about carrying out a QR factorization in connection with (3.5) applies here to the initial QR factorization (for each k where $\rho''(r(x^k))$ has one or more components near zero).

Let us now consider the new trust-region radius δ^{k+1}. Various choices for δ^{k+1} have been suggested, such as those in [Pow70b,d], [Heb73], [Mor78], [DenGW79]. They generally have the form $\delta^{k+1} = \mu^k \delta^k$ or $\delta^{k+1} = \mu^k \|D s^k\|$, and the decision whether to choose $\mu^k < 1$ or $\mu^k \geq 1$ is generally based on whether

$$\varphi(x^k) - \varphi(x^k + s^k) < c_1 [q^k(0) - q^k(s^k)] \tag{3.12}$$

for some fixed c_1. (Usually, $c_1 = 0.25$ or 0.1, and sometimes $<$ is changed to $\leq$ in (3.12).) When (3.12) holds, Moré [Mor78] chooses

$$\mu^k = \max\{0.1, \min\{0.5, \theta^*\}\} , \tag{3.13}$$

where θ^* minimizes the quadratic polynomial that fits $\gamma(0)$, $\gamma'(0)$, and $\gamma(1)$ for

$$\gamma(\theta) = \varphi(x^k + \theta s^k) , \tag{3.14}$$

i.e., $\theta^* = \frac{1}{2} \gamma'(0)/[\gamma'(0) + \gamma(0) - \gamma(1)]$ (for $\gamma(1) \neq \gamma(0) + \gamma'(0)$). (This idea for θ^* stems from [Fle71].) Hebden [Heb73] also uses (3.13), but his θ^* minimizes the cubic polynomial that fits $\gamma(0)$, $\gamma'(0)$, $\gamma''(0)$, and $\gamma(1)$, i.e. $\theta^* = [-\gamma''(0) + \sqrt{(\gamma''(0))^2 + 2\eta\gamma'(0)}]/\eta$, where $\eta = 6[\varphi(x^k + s^k) - q^k(s^k)]$. Both schemes seem

reasonable for the problems at hand. For cases where (3.12) fails, many choices for μ^k have been proposed. A new one in the spirit of (3.13) is

$$\mu^k = \max\{1, \min\{4, \theta^*\}\} \quad , \tag{3.15}$$

where θ^* has one of the values just described. The relative merit of (3.15) remains to be seen.

In cases where $\lambda^k > 0$, it may be reasonable to replace (3.14) by

$$\gamma(\theta) = \varphi(x^k + s(\theta)) \quad , \tag{3.16}$$

where $s(\theta)$ minimizes $q^k(s)$ subject to $\|Ds\|_2 \leq \theta \|D\, s^k\|_2$. Since $s'(0) = \|D\, s^k\|_2\, D^{-2} \nabla \varphi(x^k)/\|D^{-1} \nabla \varphi(x^k)\|_2$, the γ of (3.16) has $\gamma'(0) = -\|D^{-1} \nabla \varphi(x^k)\|_2 \cdot \|D\, s^k\|_2$, and since $s''(0)^T \nabla \varphi(x^k) = 0$, $\gamma''(0) = \|D\, s^k\|_2^2\, [\nabla\varphi(x^k)^T D^{-2} \nabla^2 \varphi(x^k) \cdot D^{-2} \nabla \varphi(x^k)]/\|D^{-1} \nabla \varphi(x^k)\|_2^2$. Note that $\gamma(0) = \varphi(x^k)$ and $\gamma(1) = \varphi(x^k + s^k)$ are readily available.

In addition to working well in practice, choices for δ^{k+1} of the sort just discussed make it possible to prove convergence theorems. Indeed, it is easily shown that if $H^k = \nabla^2 \varphi(x^k)$ and some iterate x^k comes close enough to a strong local minimizer x^*, i.e., a point where $\nabla\varphi(x^*) = 0$ and $\nabla^2 \varphi(x^*)$ is positive definite, then the iteration soon reduces to Newton's method and the iterates converge Q-quadratically to x^*.

Let us briefly consider what can be said more generally about convergence. Powell has studied a large class of model/trust-region algorithms in which $c_2\, \delta^k \leq \delta^{k+1} \leq c_3\, \delta^k$ when (3.12) holds and $\delta^k \leq \delta^{k+1} \leq c_4\, \delta^k$ otherwise (with $0 < c_2 \leq c_3 < 1$ and $c_4 \geq 1$), and in which H^k can be any symmetric matrix in $\mathbb{R}^{p \times p}$ (such as $\nabla^2 \varphi(x^k)$ or an approximation thereto produced by a secant update) that satisfies the bounded deterioration condition $\|H^k\| \leq c_5 + c_6 \sum_{i=0}^{k} \|s^i\|$. If $\delta^k \leq \delta^{max}$ for all k and if φ is bounded and uniformly continuous on $\{x \in \mathbb{R}^p : \|x - y\| \leq \delta^{max}$ for some y with $\varphi(y) \leq \varphi(x^k)\}$ (which is easily seen

to be the case for (2.1-6)), then Theorem 1 of [Pow75] asserts that

$$\liminf_{k \to \infty} \|\nabla\varphi(x^k)\| = 0 \quad . \tag{3.17}$$

If the rule for updating x^k is changed so that $x^{k+1} = x^k$ if (3.12) holds and $x^{k+1} = x^k + s^k$ otherwise, then (3.17) is easily strengthened to $\lim_{k\to\infty} \nabla\varphi(x^k) = 0$. Let X^* denote the set of limit points of the sequence $x^0, x^1, x^2, \cdots$ generated using the modified x^k updating rule. Continuity of $\nabla\varphi$ implies $\nabla\varphi(x^*) = 0$ for all $x^* \in X^*$, and if X^* contains just one point x^*, then $\lim_{k\to\infty} x^k = x^*$. If $H^k = \nabla^2 \varphi(x^k)$ and X^* contains more than one point, then $\nabla^2 \varphi(x^*)$ is positive semidefinite and singular at each $x^* \in X^*$. Although pathological cases where X^* contains two or more points are extremely unlikely in practice, I know of no easy way to exclude them in theory. So far as the rule for updating x^k is concerned, it may be better in practice to set $x^{k+1} = x^k + s^k$ whenever $\varphi(x^k + s^k) < \varphi(x^k)$, since (in our experience) this leads to faster convergence more often than not.

When iterative methods are used to solve a problem, the overall time required to find an acceptable solution may be less for a more slowly convergent method than for a more rapidly convergent one that requires more work per iteration. One way to reduce the work per iteration in the above model/trust-region approach is by sometimes choosing H^k to be H^{k-1} or a cheaply updated version of H^{k-1}. If H^k is periodically chosen to be $\nabla^2 \varphi(x^k)$ and is otherwise set equal to H^{k-1} (which Traub suggested in §11.3 of [Tra64]), then it is possible to use Brent's ideas [Bre73] to optimize the asymptotic efficiency of the iteration once it has reduced to (3.6) with $\lambda_k = 1$. In practice, it might be better to use an idea of Todd [SaiT78]: periodically choose H^k to be $\nabla^2 \varphi(x^k)$ and otherwise obtain H^k by performing a secant update on H^{k-1}. A similar idea is possible with H^k given periodically by (3.8). At this point it is not clear just which combination of the ideas presented in this section will minimize the time needed to find an acceptable local minimizer of (1.1) when $r(x)$ has the form (3.1).

4. Nonlinear Problems

In this section we generalize the discussion in §3 to the case where the residual vector $r(x)$ is a twice differentiable function of x. In place of (3.2) and (3.3), we then have

$$\nabla\varphi(x) = J(x)^T \rho'(r(x)) \quad \text{and} \tag{4.1}$$

$$\nabla^2 \varphi(x) = G(x) + S(x) \quad \text{where} \tag{4.2a}$$

$$G(x) = J(x)^T D(\rho''(r(x)))J(x) \quad \text{and} \tag{4.2b}$$

$$S(x) = \sum_{i=1}^{n} \rho_i'(r_i(x))\nabla^2 r_i(x) \quad . \tag{4.2c}$$

Aside from direct search algorithms (i.e., pattern searchers, such as the method of simplices [NelM65]), most minimization algorithms require a good approximation to the gradient of the objective function. It is usually possible to obtain an acceptable approximation by finite differences. For the objective function φ of present concern, it is slightly cheaper to compute a finite-difference approximation $J(x, h)$ to $J(x)$ and then compute the approximate gradient

$$\nabla\varphi(x, h) = J(x, h)^T \rho'(r(x)) \tag{4.3}$$

than it is to approximate $\nabla\varphi(x)$ directly by finite differences, since the criterion functions need only be evaluated at $r(x)$ when (4.3) is used. Thus it is reasonable to assume that a good approximation to $J(x)$ is available, and to simplify the following notation and discussion, we henceforth assume that $J(x)$ itself is available.

Because of $S(x)$ in (4.2), algorithms explicitly using $\nabla^2 \varphi(x)$ would be difficult to implement and expensive to run. On problems where $S(x)$ is likely to be small at the desired solution, it is reasonable to discard $S(x)$ and thus replace $\nabla^2 \varphi(x^k)$ by its Gauss-Newton approximation $H^k = G(x^k)$. Using this H^k in the methods of §3, we obtain the Gauss-Newton analogs of these methods.

For many nonlinear problems, $G(x^k)$ gives only a poor approximation to $\nabla^2 \varphi(x^k)$, and better performance may well be obtained if H^k is generated by a

secant update method [DenM77]. However, since the Gauss-Newton part $G(x^k)$ of the true Hessian $\nabla^2 \varphi(x^k)$ is readily available, it seems reasonable to approximate only the expensive part $S(x^k)$ of (4.2) by a secant update scheme. This works well in the particular case of nonlinear least-squares [DenGW79], and it is natural to expect that the obvious generalizations sketched below of the techniques recommended in [DenW78] and [DenGW79] may prove worthwhile for other problems of the form (1.1).

Let us first consider how to update a current approximation S^k to $S(x^k)$. Suppose step s^k has been taken (i.e., $x^{k+1} = x^k + s^k$). As in [DenGW79], it seems reasonable to require that $S^{k+1} s^k \doteq S(x^{k+1})s^k$, and

$$S(x^{k+1})s^k = \sum_{i=1}^{n} \rho_i'(r_i(x^{k+1}))\nabla^2 r_i(x^{k+1})s^k$$

$$\doteq \sum_{i=1}^{n} \rho_i'(r_i(x^{k+1}))[\nabla r_i(x^{k+1}) - \nabla r_i(x^k)]$$

$$= \left[J(x^{k+1}) - J(x^k)\right]^T \rho'(r(x^{k+1})) \quad ,$$

so we shall require

$$S^{k+1} s^k = y^k := \left[J(x^{k+1}) - J(x^k)\right]^T \rho'(r(x^{k+1})) \quad . \tag{4.4}$$

While there are many ways of obtaining an S^{k+1} that satisfies (4.4), the reasoning in §3 of [DenGW79] and in [Gay76] suggests using

$$\Delta g^k := \nabla\varphi(x^{k+1}) - \nabla\varphi(x^k) \tag{4.5a}$$

$$v^k := \Delta g^k / [(s^k)^T \Delta g^k] \tag{4.5b}$$

$$z^k := y^k - S^k s^k \tag{4.5c}$$

$$w^k := z^k - \frac{1}{2}[(z^k)^T s^k]v^k \tag{4.5d}$$

$$S^{k+1} := S^k + v^k(w^k)^T + w^k(v^k)^T \quad , \tag{4.5e}$$

at least when $(s^k)^T \Delta g^k > 0$. To handle cases where s^k and Δg^k are nearly orthogonal, it seems reasonable to replace (4.5b) by

$$v^k := \begin{cases} 0 & \text{if } \Delta g^k = 0 \\ \Delta g^k / [\operatorname{sign}((s^k)^T \Delta g^k) \max\{|(s^k)^T \Delta g^k|, c_5 \|s^k\|_2 \, \|\Delta g^k\|_2] & \text{otherwise,} \end{cases} \tag{4.6}$$

where c_5 is a small positive number (such as 10^{-4} or 10^{-6}).

For the nonlinear least-squares problem, it proved worthwhile to <u>size</u> S^k before updating it, i.e., to replace S^k in (4.5) by $\min\{1, |\tau^k|\}S^k$, where

$$\tau^k = (y^k)^T s^k / [(s^k)^T S^k s^k] \quad . \tag{4.7}$$

Multiplying S^k by τ^k shifts its spectrum (interval from minimum to maximum eigenvalue), making it more likely to overlap that of $S(x^{k+1})$, which has the happy effect of making S^{k+1} small in cases where $S(x^{k+1})$ is small.

While working with nonlinear least-squares problems, we tried several other candidates for τ^k, including Welsch's proposal in [DenW78], and concluded that (4.7) looked best on the problems considered. Whether the same conclusion holds for the more general problems of interest here remains to be seen.

Our experience with nonlinear least-squares suggests using a model/trust-region approach in which there are two models: the <u>Gauss-Newton model</u>, in which $H^k = G(x^k)$, and the <u>augmented model</u>, in which $H^k = G(x^k) + S^k$. To decide which model to use in determining x^{k+2}, we found the following rule adequate: if

$$|q^k(s^k) - \varphi(x^k + s^k)| \leq c_6 |\dot{q}^k(s^k) - \varphi(x^k + s^k)| \quad , \tag{4.8}$$

where q^k is the current model and $\dot{q}^k$ is the alternate model, then retain the same model preference; otherwise switch models. (We found $c_6 \doteq \sqrt{2}$ to work well in (4.8), though we did not experiment much with this constant.) This is called <u>adaptive modeling</u>.

An important practical detail is the matter of deciding when an acceptable solution has been found. The discussion in §6 of [DenGW79] generalizes readily to

objective functions of the form (1.1). Specifically, if $\nabla\varphi(x^*) = 0$ with $\rho'(r(x^*)) \neq 0$, then (4.1) implies that $\rho'(r(x^*))$ is orthogonal to the columns of $J(x^*)$, which suggests checking for cosine convergence, i.e.,

$$(4.9) \qquad \max\left\{\frac{|\rho'(r(x^k))^T J_i(x^k)|}{\|\rho'(r(x^k))\|_2 \, \|J_i(x^k)\|_2} : \|J_i(x^k)\|_2 > \varepsilon_{Ji} \ , \ 1 \le i \le p\right\} \le \varepsilon_c \ ,$$

where J_i denotes column i of J and $\varepsilon_{Ji} \ge 0$ is a tolerance used to decide whether J_i should be regarded as a zero vector. Except for the choice of the ε_{Ji}, test (4.9) has the advantage of being unaffected by how the components of x are scaled. On problems where $\rho'(r(x^*)) = 0$, a test of the form (4.9) may fail no matter how close x^k is to x^*, so it is also necessary to check for generalized residual convergence, i.e.

$$(4.10) \qquad \|\rho'(r(x^k))\| \le \varepsilon_R \ ,$$

which unfortunately is sensitive to the scale of $r(x)$. To handle cases where ε_c and ε_R are too small for the precision of the arithmetic being used, it is advisable to check for X convergence - whether s^k is small relative to x^k when s^k is rejected, i.e., when $x^{k+1} = x^k$. One way to do this is by checking whether

$$(4.11) \qquad \max\left\{\frac{|\mathrm{fl}(x_i^k + s_i^k) - x_i^k|}{|x_i^k + s_i^k| + |x_i^k|} : 1 \le i \le p\right\} \le \varepsilon_X \ ,$$

where $\mathrm{fl}(\cdot)$ denotes the computed value of $(\cdot)$. This X convergence test is independent of the scale of the components of x^k and s^k, but may have trouble if $x_i^k \doteq 0$ for some i. We can, of course, replace the componentwise test (4.11) by one of the form

$$(4.12) \qquad \|D[\mathrm{fl}(x^k + s^k) - x^k]\| \le \varepsilon_X \|D\, x^k\| \ ,$$

which is scale-invariant if the scale matrix D is chosen properly and only fails to perform properly if $x^* \doteq 0$. (We may wish to change D from one iteration to the next - see §7 of [DenGW79].) It may also be reasonable to consider a generalization

of the variability convergence test described in [DenGW79]. Although it may not always be clear how to generalize the scale factor $\sum_{i=1}^{n} r_i(x^k)^2/\max\{1, n - p\}$ used for nonlinear least-squares, it is clear that the alternate interpretation of the variability convergence test in the case of a full Newton step generalizes to a test on the predicted change yet possible in the objective function. To handle cases where s^k is not a Newton step (i.e., $H^k s^k \neq -\nabla\varphi(x^k)$), it may be best to check whether, say,

$$\varphi(x^k) - \min\{q^k(s^k)\ ,\ \varphi(x^k + s^k)\} \leq \varepsilon_F \cdot |\varphi(x^k)|\ , \tag{4.13}$$

which if true, might be called function convergence. For reasonable values of ε_F, it is likely that (4.13) would be satisfied sooner than (4.9-12).

Acknowledgment

I thank Roy E. Welsch for some helpful discussions and for useful comments on a draft of this paper.

Appendix A: Computing $(D^{1/2} \bar{A}) \backslash (D^{-1/2} \bar{b})$

In this appendix we consider in more detail a problem in numerical linear algebra from §3 : given $\bar{A} \in \mathbb{R}^{n \times p}$ of rank p, $\bar{b} \in \mathbb{R}^n$, and $D = \text{diag}\,(d_1, \cdots, d_n)$ with $d_i > 0$, $1 \le i \le n$, compute a good approximation s to

$$s^* = (\bar{A}^T D \bar{A})^{-1} \bar{A}^T \bar{b} \tag{A.1}$$

using finite-precision arithmetic. Let

$$A = D^{1/2} \bar{A} \quad \text{and} \tag{A.2a}$$

$$b = D^{-1/2} \bar{b} \ , \tag{A.2b}$$

so that (A.1) amounts to $s^* = (A^T A)^{-1} A^T b$. How easily we can compute a good s depends largely on the condition number κ of A, i.e., the ratio of largest to smallest singular value of A, which is given by

$$\kappa = \left(\|A^T A\|_2 \, \|(A^T A)^{-1}\|_2 \right)^{1/2} \ . \tag{A.3}$$

Let $\text{fl}(\cdot)$ denote the result of computing $(\cdot)$ in the available finite-precision arithmetic. We assume this to be floating-point arithmetic, so that if $\underline{\text{op}}$ denotes one of the four elementary arithmetic operations and $\alpha \,\underline{\text{op}}\, \beta$ is defined and in range, then $\text{fl}(\alpha \,\underline{\text{op}}\, \beta) = (\alpha \,\underline{\text{op}}\, \beta)(1 + \eta)$ for some η with $|\eta| \le \varepsilon_{\text{MACH}}$, where $\varepsilon_{\text{MACH}}$ (the "machine epsilon") depends only on the arithmetic being used. (If binary floating-point arithmetic having t bits in the fraction is used, then $\varepsilon_{\text{MACH}}$ is generally 2^{-t} or 2^{1-t}, depending on whether results are rounded or truncated. It is often satisfactory to define $\varepsilon_{\text{MACH}}$ as the smallest positive floating-point number such that $\text{fl}(1 + \varepsilon_{\text{MACH}}) > 1$ and $\text{fl}(1 - \varepsilon_{\text{MACH}}) < 1$.)

If κ is sufficiently less than $\varepsilon_{\text{MACH}}^{-1/2}$, then the most efficient way to compute an acceptable approximation s to s^* is to explicitly compute $\text{fl}(\bar{A}^T D \bar{A})$ and solve $\text{fl}(\bar{A}^T D \bar{A})s = \text{fl}(\bar{A}^T \bar{b})$ by computing a Cholesky factorization of $\text{fl}(\bar{A}^T D \bar{A})$. A roundoff analysis of this procedure leads to a relative error bound analogous to (A.10) below (provided that $\kappa^2 \varepsilon_{\text{MACH}}$ is not too large).

Unfortunately, people often overspecify their models, which can easily lead to $\kappa \geq \varepsilon_{MACH}^{-1/2}$. In this case it is possible for $fl(\bar{A}^T D \bar{A})$ to be indefinite or singular, and the procedure just sketched can break down or yield a poor s. If $\varepsilon_{MACH}^{-1/2} \leq \kappa < \varepsilon_{MACH}^{-1}$, then, as indicated in §3, it is usually possible to obtain a much better s by properly computing a QR factorization of $A = D^{1/2} \bar{A}$ and approximately solving $Rs = Q^T b$. We now consider how to properly compute a QR factorization in cases where some d_i may be near zero.

In general, if $A \in \mathbb{R}^{n \times p}$ and the columns of A are readily available, it is common practice to compute a QR factorization of A by means of Householder transformations (elementary reflectors). In this case Q amounts to the first p columns of the product $Q^1 Q^2 \cdots Q^p$, where $Q^k = I - 2(\|u^k\|_2^2)^\dagger u^k (u^k)^T$, $\tau^\dagger = \begin{cases} \tau^{-1} & \text{if } \tau \neq 0 \\ 0 & \text{if } \tau = 0 \end{cases}$ for $\tau \in \mathbb{R}$, and $u^k \in \mathbb{R}^n$ is chosen so that $A^{k+1} := Q^k \cdots Q^1 A$ has $A^{k+1}_{i,j} = 0$ for $i > j$, $1 \leq j \leq k$. Starting from $A^1 := A$, this is accomplished by choosing

$$u^k_i = \begin{cases} 0 & \text{if } i < k \\ A^k_{k,k} + \operatorname{sign}(A^k_{k,k}) \left[\sum_{j=k}^{n} (A^k_{j,k})^2 \right]^{1/2} & \text{if } i = k \\ A^k_{i,k} & \text{if } i > k \ . \end{cases}$$

With this scheme, $Q^T b$ is the first p components of the b^{p+1} computed by $b^1 := b$, $b^{k+1} := b^k - 2(\|u^k\|_2^2)^\dagger [(u^k)^T b^k] u^k$ for $1 \leq k \leq p$. Unfortunately, if $d_k \doteq 0$ for some $k < p$, then this scheme can lead to severe cancellation errors in components k through p of $fl(Q^p \cdots Q^{k+1} b^k)$. Suppose for example that

$\bar{A} = \begin{pmatrix} 1 & 1 \\ .01 & 2 \\ 1 & 1 \end{pmatrix}$, $\bar{b} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$, and $D = \operatorname{diag}(.01, 1, 1)$, so that

$$A = \begin{pmatrix} .01 & .01 \\ .01 & 2 \\ 1 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} 100 \\ 1 \\ 2 \end{pmatrix} . \tag{A.4}$$

If we use 3 decimal rounded floating-point arithmetic ($\varepsilon_{MACH} = .005$) and obtain s from a QR factorization of A computed as above (using division by $fl(u_k^k \cdot A_{k,k}^k)$ in place of multiplication by $2\|u^k\|_2^{-2} = (u_k^k \cdot A_{k,k}^k)^{-1}$), then we get $s = \begin{pmatrix} 3.53 \\ -.518 \end{pmatrix}$, whereas $s^* \doteq \begin{pmatrix} 2.51226 \\ .487439 \end{pmatrix}$. If the components of A, b, and d had been ordered to make d_3 the small one, then the computed s would have been much more accurate. For example, if we interchange rows 1 and 3, so that

$$A = \begin{pmatrix} 1 & 1 \\ .01 & 2 \\ .01 & .01 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 2 \\ 1 \\ 100 \end{pmatrix}, \tag{A.5}$$

then we obtain $s = \begin{pmatrix} 2.52 \\ .490 \end{pmatrix}$. This illustrates a way of avoiding disasterous cancellation errors when using Householder transformations: we cound introduce row pivoting, i.e., before generating Q^k, we could arrange that $|A_{k,k}^k| = \max\{|A_{i,k}^k| : k \le i \le n\}$ by interchanging two rows of $[A^k \vdots b^k]$ if necessary. (Equivantly, we could use a permutation vector in the obvious way.)

A QR factorization computed using Givens transformations (plane rotations) can also suffer disasterous cancellation errors in some cases. With this scheme, Q is the first p columns of a product of matrices $Q^{k,\ell}$ of the form

$$Q_{i,j}^{k,\ell} = \begin{cases} \cos \theta^{k,\ell} & \text{if } i = j = k \text{ or } i = j = \ell \\ \sin \theta^{k,\ell} & \text{if } i = k \text{ and } j = \ell \\ -\sin \theta^{k,\ell} & \text{if } i = \ell \text{ and } j = k \\ 1 & \text{if } \ell \neq i = j \neq k \\ 0 & \text{otherwise.} \end{cases}$$

If A and b are given by (A.4), for example, and we compute s using the Givens transformations $Q^{1,2}\, Q^{1,3}\, Q^{2,3}$ computed in the same 3 decimal arithmetic as before, then we obtain $s = \begin{pmatrix} 2.35 \\ .650 \end{pmatrix}$. The substantial error in this s comes about because

$\vartheta^{1,2}$ is far from $\pm \pi/2$, which results in a significant cancellation error in $\mathrm{fl}(Q^T b)$. Here again appropriate row pivoting can greatly reduce the critical cancellation errors. For instance, if A and b are given by (A.5), then we obtain $s = \begin{pmatrix} 2.51 \\ .487 \end{pmatrix}$.

The stabilized Gram-Schmidt process gives good performance without any row pivoting. In this case we compute a QR factorization of A by the following algorithm, in which q^k denotes the kth column of Q and $A^k_{\cdot,j}$ denotes column j of A^k.

$A^1 := A$

For $k = 1, 2, \cdots, p$:

$R_{k,k} := \|A^k_{\cdot,k}\|_2$

$q^k := A^k_{\cdot,k} / R_{k,k}$

For $j = k + 1, \cdots, p$:

$R_{k,j} := (q^k)^T A^k_{\cdot,j}$

$A^{k+1}_{\cdot,j} := A^k_{\cdot,j} - R_{k,j} q^k$

On both (A.4) and (A.5) this leads to $s = \begin{pmatrix} 2.51 \\ .487 \end{pmatrix}$ when 3 deciman arithmetic is used as before.

We may use floating-point error analysis to justify the above claim about good performance from the stabilized Gram-Schmidt process. It will be useful to introduce the notation $O_{n,p}(\tau)$ to denote a quantity such that $0 \le O_{n,p}(\tau) \le \tau \cdot \zeta(n, p)$ for some low order polynomial $\zeta(n, p)$ and all $\tau \ge 0$. In this notation, the conventional big O is $O(\tau) := O_{1,1}(\tau)$.

When we compute a QR decomposition of A and an approximation v to $Q^T b$ using one of the schemes sketched above, it is generally possible to show that

(A.6a) $$A + E = QR \quad \text{and}$$

(A.6b) $$v + f = Q^T b \ ,$$

where $\|E\| = O_{n,p}(\varepsilon_{MACH} \cdot \|A\|)$ and $\|f\| = O_{n,p}(\varepsilon_{MACH} \cdot \|b\|)$, provided $n \cdot \varepsilon_{MACH}$ is small (e.g. $n \cdot \varepsilon_{MACH} \leq .01$). Wilkinson [Wil65] does this for Householder transformations, Gentleman [Gen73] does this for (standard and fast) Givens transformations, and Björck [Bjö67] does this for the stabilized Gram-Schmidt process. Relatively little error occurs when we compute s by solving $Rs = v$; we could take this small error into account by adjusting E and f, but doing so would not change the character of the bounds on E and f, so we shall assume this error to be zero. Another source of error with the stabilized Gram-Schmidt process is the fact that the columns of the computed Q are usually not truly orthonormal. Since this fact also does not change the nature of the perturbation bound (A.10) given below (see [Bjö67]), we shall ignore it too. Thus we consider that the computed s exactly solves a modified problem: $s = (A + E)^{\dagger}(b - Qf)$, where $(A + E)^{\dagger} = [(A + E)^T(A + E)]^{-1}(A+E)^T$ denotes the pseudoinverse of $A + E$, which, like A, we assume to have rank p.

In discussing the difference between the computed s and the desired one, i.e., $s - s^* = (A + E)^{\dagger}(b - Qf) - A^{\dagger} b$, it is useful to use the notation P_A for orthogonal projection onto the column space of A : $P_A = A A^{\dagger} = A(A^T A)^{-1} A^T$. In this notation, $I - P_A$ denotes (orthogonal) projection onto the orthogonal complement of the column space of A.

When A and b come from a general linear least-squares problem, $\|(I - P_A)b\|/\|P_A b\|$ is often not too large, in which case the perturbation bounds that Stewart derives in [Ste69] are quite satisfactory: Theorem 6.2 of [Ste69] shows for, say, $\kappa \|P_A E\|/\|A\| \leq 1/2$ that

$$\text{(A.7)} \quad \frac{\|(A+E)^{\dagger} b - A^{\dagger} b\|}{\|A^{\dagger} b\|} = O\left(\kappa \cdot \frac{\|P_A E\|}{\|A\|} + \kappa^2 \cdot \frac{\|(I - P_A)b\|}{\|P_A b\|} \cdot \frac{\|(I - P_A)E\|}{\|A\|} + \kappa^3 \cdot \frac{\|(I - P_A)E\|^2}{\|A\|^2}\right) \ ,$$

where κ is given by (A.3). When A and b come from (A.2), however,

$\|(I - P_A)b\|/\|P_A\, b\|$ and therewith the right-hand side of (A.7) can be arbitrarily large. Fortunately, the left-hand side of (A.7) only depends on A, E, $A^T b$, and $E^T b$. Indeed, we prove in Appendix B that if, say, $\kappa\, \|P_A E\|/\|A\| \le 1/2$, then

$$(A.8)\qquad \frac{\|(A+E)^{+} b - A^{+} b\|}{\|A^{+} b\|} = O\left(\kappa^2 \cdot \left[\frac{\|E\|}{\|A\|} + \frac{\|E^T b\|}{\|A\|\,\|P_A\, b\|}\right] + \left[\kappa^2 \frac{\|E\|}{\|A\|}\right]^2 \frac{\|(A+E)^T b\|}{\|A\|\,\|P_A\, b\|}\right) .$$

It is straightforward to modify the derivations in [Bjö67] to show that

$$(A.9a)\qquad \|E^T b\| = O_{n,p}(\varepsilon_{MACH} \cdot a^T|b|) \quad \text{and}$$

$$(A.9b)\qquad \|f\| = O_{n,p}(\varepsilon_{MACH} \cdot \kappa \cdot a^T|b|/\|A\|)$$

for the stabilized Gram-Schmidt process, where $a \in \mathbb{R}^n$ has ith component $a_i = \max\{|A_{i,j}| : 1 \le j \le p\}$ and $|b| \in \mathbb{R}^n$ has ith component $|b_i|$. Together with (A.6) and (A.8), this may be used to show that

$$(A.10)\qquad \frac{\|(A+E)^{+}(b - Qf) - A^{+} b\|}{\|A^{+} b\|} = O_{n,p}\left(\kappa^2 \cdot \varepsilon_{MACH} \cdot \left[\frac{a^T|b|}{\|A\|\,\|P_A\, b\|} + 1\right]\right)$$

(for, say, $\kappa \cdot \varepsilon_{MACH} \cdot \zeta(n, p) \le 1/2$).

While I suspect that (A.9) usually holds when Householder or Givens transformations with row pivoting are used to compute a QR factorization, at this writing I am unable to obtain satisfactory bounds for these procedures.

Appendix B: Bounding $(A + E)^{\dagger} b - A^{\dagger} b$

In this appendix we may gain some generality by allowing A, E, and b to have complex components. Thus we assume that $b \in \mathbb{C}^{n\times p}$, $E \in \mathbb{C}^{n\times p}$, and that $A \in \mathbb{C}^{n\times p}$ has rank p. Superscript H stands for conjugate transpose, and $\|\cdot\|$ denotes the Euclidean vector norm $\|x\| = \|x\|_2 := (x^H x)^{1/2}$ or the corresponding induced matrix norm: $\|A\| := \max\{\|Ax\| : \|x\| = 1\}$. As before, κ denotes the condition number of A, i.e., $\kappa = [\|A^H A\| \, \|(A^H A)^{-1}\|]^{1/2}$, $A^{\dagger} = (A^H A)^{-1} A^H$ is the pseudoinverse of A, and $P_A = A A^{\dagger}$ denotes the operator that projects orthogonally onto the column space of A. It will be convenient to use the notation

$$\beta := \kappa \|E\|/\|A\| \tag{B.1a}$$

$$\beta_1 := \kappa \|P_A E\|/\|A\| \quad . \tag{B.1b}$$

It will also be convenient to exploit the fact that A has a QR decomposition: $A = QR$, where $Q \in \mathbb{C}^{n\times p}$ has orthonormal columns (i.e., $Q^H Q = I$) and $R \in \mathbb{C}^{p\times p}$ is upper triangular with $\|R\| = \|A\|$, $\|R^{-1}\| = \|A^{\dagger}\| = \kappa/\|A\|$. Note that $P_A = Q Q^H$.

Theorem B.3 below rests heavily on Lemma B.2, which in turn relies on the following simple lemma:

Lemma B.1 Assume $F \in \mathbb{C}^{n\times p}$ and let $F_1 = P_A F$. If $\|F_1\| < 1$, then $(Q + F)^H (Q + F)$ is nonsingular and

$$\|I - Q^H F - F^H Q - [(Q + F)^H (Q + F)]^{-1}\| \le \frac{4\|F_1\|^2 + \|F\|^2 [1 + 2\|F_1\|]}{(1 - \|F_1\|)^2} \quad . \tag{B.2}$$

Proof: For any $x \in \mathbb{C}^p$,

$$\|(Q + F)x\| \ge \|Q^H (Q + F)x\| = \|(I + Q^H F_1)x\| \ge (1 - \|F_1\|)\|x\| \quad ,$$

so the smallest eigenvalue of $(Q + F)^H (Q + F)$ is at least $(1 - \|F_1\|)^2$, whence $(Q + F)^H (Q + F)$ is nonsingular and

$$\|[(Q + F)^H (Q + F)]^{-1}\| \le (1 - \|F_1\|)^{-2} \quad . \tag{B.3}$$

Now $[(Q + F)^H(Q + F)][I - Q^H F - F^H Q] = I - (Q^H F + F^H Q)(Q^H F + F^H Q) +$

$+ F^H F(I - Q^H F - F^H Q)$, so

$$I - Q^H F - F^H Q - [(Q + F)^H(Q + F)]^{-1} =$$

(B.4)

$$= [(Q + F)^H(Q + F)]^{-1}[-(Q^H F + F^H Q)(Q^H F + F^H Q) + F^H F(I - Q^H F - F^H Q)] \quad ,$$

and (B.2) follows readily from (B.3), (B.4), and the fact that $\|Q^H F\| = \|F^H Q\| = \|F_1\|$. ∎

Lemma B.2 Let $F = E R^{-1}$, $F_1 = P_A F$, and assume $\beta_1 < 1$. Then $\|F_1\| < 1$, $(A + E)^H(A + E)$ is nonsingular, and if P_{A+E} is the orthogonal projection onto the column space of $A + E$, i.e., $P_{A+E} = (A + E)(A + E)^+$, then

$$\|P_A(P_{A+E} - P_A)b\| \leq \|F^H(I - Q Q^H)b\| + \|F_1\|\,\|F^H b\| + [2\|F\|\,\|F_1\| +$$

$$+ \frac{(1 + \|F\|)(4\|F_1\|^2 + \|F\|^2(1 + 2\|F_1\|))}{(1 - \|F_1\|)^2}]\,\|(Q + F)^H b\|$$

(B.5)

$$\leq \frac{\kappa}{\|A\|}\,[\|E^H(I - P_A)b\| + \beta_1\,\|E^H b\| +$$

$$+ (2\beta\beta_1 + \frac{[1 + \beta][4\beta_1^2 + \beta^2 + 2\beta_1\beta^2]}{[1 - \beta_1]^2})\,\|(A + E)^H b\|] \quad .$$

Proof: We have $\|F_1\| = \|P_A E R^{-1}\| \leq \|R^{-1}\|\,\|P_A E\| \leq \kappa\,\|P_A E\|/\|A\| = \beta_1$, so $(Q + F)^H(Q + F)$ and $(A + E)^H(A + E) = R^H(Q + F)^H(Q + F)R$ are nonsingular by Lemma B.1.

Let $M = [(Q + F)^H(Q + F)]^{-1} - [I - Q^H F - F^H Q]$. Since $P_{A+E} = (Q + F)[(Q + F)^H(Q + F)]^{-1}(Q + F)^H$, we have

$$P_{A+E} - P_A = (Q + F)[I - Q^H F - F^H Q](Q + F)^H - Q Q^H + (Q + F)M(Q + F)^H$$

(B.6)

$$= [I - Q Q^H]F(Q + F)^H + Q F^H(I - Q Q^H) - Q F^H Q F^H +$$

$$+ [(Q + F)M - F(Q^H F + F^H Q)](Q + F)^H \quad .$$

Now $\|Q\| = 1$, $P_A = Q Q^H$, and $P_A(I - P_A) = 0$, so the first inequality in (B.5)

follows from (B.6) and Lemma B.1 (which bounds $\|M\|$). As above, we have $\|F_1\| \leq \beta_1$ and similarly $\|F\| \leq \kappa \|E\|/\|A\| = \beta$ (see (B.1)); because of these inequalities and the facts that $\|F^H(I - Q Q^H)b\| \leq \kappa \|E^H(I - P_A)b\|/\|A\|$ and $\|(Q + F)^H b\| \leq$ $\leq \kappa \|(A + E)^H b\|/\|A\|$, the second inequality in (B.5) follows from the first. ∎

We may now prove the main result of this appendix:

Theorem B.3: If $\beta_1 < 1$, then

$$\frac{\|(A + E)^{\dagger} b - A^{\dagger} b\|}{\|A^{\dagger} b\|} \leq \left(\frac{\kappa}{1 - \beta_1}\right)\left(\frac{\|P_A E\| + \kappa\|E\|}{\|A\|} + \frac{\kappa}{\|A\| \, \|P_A b\|}\left[(1 + \beta_1)\|E^H b\| + \right.\right.$$

(B.7)

$$\left.\left. + (2\beta\beta_1 + \frac{[1 + \beta][4\beta_1^2 + \beta^2 + 2\beta^2\beta_1]}{(1 - \beta_1)^2})\|(A + E)^H b\|\right]\right) .$$

Proof: From (4.8) of [Ste69] we have

$$(B.8) \qquad (A + E)^{\dagger} b - A^{\dagger} b = (A + P_A E)^{\dagger}[P_A(P_{A+E} - P_A)b + P_A E A^{\dagger} b] .$$

By (6.5) of [Ste69], $\|(A + P_A E)^{\dagger}\| \leq \|A^{\dagger}\|/(1 - \beta_1)$. This combines with (B.8), Lemma B.2, and the facts that $\kappa = \|A\| \, \|A^{\dagger}\|$, $P_A = A A^{\dagger}$, and $\|A^{\dagger} b\| \geq \|P_A b\|/\|A\|$ to give (B.7). ∎

REFERENCES

[Beat74] Beaton, A. E. and Tukey, J. W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data", Technometrics 16, pp. 147-185.

[Bjö67] Björck, Å. (1967), "Solving Linear Least Squares Problems by Gram-Schmidt Orthogonalization", BIT 7, pp. 1-21.

[Bre73] Brent, R. P. (1973), "Some Efficient Algorithms for Solving Systems of Nonlinear Equations", SIAM J. Numer. Anal. 10, pp. 327-344.

[BunP71] Bunch, J. R. and Parlett, B. N. (1971), "Direct Methods for Solving Symmetric Indefinite Systems of Linear Equations", SIAM J. Numer. Anal. 8, pp. 639-655.

[ByrP79] Byrd, R. H. and Pyne, D. A. (1979), "On the Convergence of Iteratively Reweighted Least Squares for Robust Regression", Technical Report, Department of Mathematical Sciences, Johns Hopkins University.

[DenGW79] Dennis, J. E., Gay, D. M. and Welsch, R. E. (1979), "An Adaptive Nonlinear Least-Squares Algorithm", Report TR 77-321 (revised), Department of Computer Science, Cornell University.

[DenM77] Dennis, J. E. and Moré, J. J. (1977), "Quasi-Newton Methods, Motivation and Theory", SIAM Rev. 19, pp. 46-89.

[DenS79] Dennis, J. E. and Schnabel, R. B. (1979), "Quasi-Newton Methods for Unconstrained Nonlinear Problems", manuscript.

[DenW78] Dennis, J. E. and Welsch, R. E. (1978), "Techniques for Nonlinear Least Squares and Robust Regression", Comm. Statist. B7, pp. 345-359.

[Fai74] Fair, R. C. (1974), "On the Robust Estimation of Econometric Models", Ann. Econom. Social Measurement 3, pp. 667-677.

[Fle71] Fletcher, R. (1971), "A Modified Marquardt Subroutine for Nonlinear Least Squares", Report R6799, A.E.R.E. Harwell, Oxon., England.

[FleF75] Fletcher, R. and Freeman, T. L. (1975), "A Modified Newton Method for Minimisation", Report No. 7 (Numerical Analysis Reports), University of Dundee, Department of Mathematics.

[Gay76] Gay, D. M. (1976), "Representing Symmetric Rank 2 Updates", NBER Working Paper No. 124.

[Gay79] Gay, D. M. (1979), "Computing Optimal Locally Constrained Steps", in preparation.

[Gen73] Gentleman, W. M. (1973), "Least Squares Computations by Givens Transformations Without Square Roots", J. Inst. Math. Appl. 12, pp. 329-336.

[GilM74] Gill, P. E. and Murray, W. (1974), "Safeguarded Steplength Algorithms for Optimization Using Descent Methods", Report NAC37, Nat'l Physical Lab., England.

[Gol77] Goldfarb, D. (1977), "Curvilinear Path Steplength Algorithms for Minimization Which Use Directions of Negative Curvature", Report CCNY-CS-77-101, Department of Computer Science, City College of City University of New York.

[GolQT66] Goldfeld, S. M., Quandt, R. E. and Trotter, H. F. (1966), "Maximization by Quadratic Hill-Climbing", Econometrica 34, pp. 541-551.

[Gre67] Greenstadt, J. (1967), "On the Relative Efficiencies of Gradient Methods", Math. Comput. 21, pp. 360-367.

[Heb73] Hebden, M. D. (1973), "An Algorithm for Minimization Using Exact Second Derivatives", report TP515, A.E.R.E. Harwell, Oxfordshire, England.

[HinT75] Hinich, M. J. and Talwar, P. P. (1975), "A Simple Method for Robust Regression", J. Amer. Statist. Assoc. 70, pp. 113-119.

[HolW77] Holland, P. W. and Welsch, R. E. (1977), "Robust Regression Using Iteratively Reweighted Least-Squares", Comm. Statist. A6, pp. 813-827.

[Hub64] Huber, P. J. (1964), "Robust Estimation of a Location Parameter", Ann. Math. Statist. 35, pp. 73-101.

[Hub75] Huber, P. J. (1975), "Robust Methods of Estimation of Regression Coefficients", manuscript for talk at Second International Summer School on Problems of Model Choice and Regression Analysis at Rheinhardsbrum, G.D.R., November 8-18.

[LawH74] Lawson, C. L. and Hanson, R. J. (1974), Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, N.J.

[Lev44] Levenberg, K. (1944), "A Method for the Solution of Certain Nonlinear Problems in Least Squares", Quart. Appl. Math. 2, pp. 164-168.

[Mar63] Marquardt, D. W. (1963), "An Algorithm for Least Squares Estimation of Nonlinear Parameters", SIAM J. Appl. Math. 11, pp. 431-441.

[McC77] McCormick, G. (1977), "A Modification of Armijo's Step-Size Rule for Negative Curvature", Math. Programming 13, pp. 111-115.

[Mor78] Moré, J. J. (1978), "The Levenberg-Marquardt Algorithm: Implementation and Theory", pp. 105-116 of Lecture Notes in Mathematics 630, edited by G. A. Watson, Springer-Verlag, Berlin, Heidelberg and New York.

[MorS77] Moré, J. J. and Sorensen, D. C. (1977), "On the Use of Directions of Negative Curvature in a Modified Newton Method", Report TM-319, Applied Math. Div., Argonne National Laboratory.

[Mur72] Murray, W. (1972), "Second Derivative Methods", pp. 57-71 of Numerical Methods for Unconstrained Optimization, edited by W. Murray, Academic Press, London and New York.

[NelM65] Nelder, J. A. and Mead, R. (1965), "A Simplex Method for Function Minimization", Comput. J. 7, pp. 308-313.

[NelW72] Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models", J. Roy. Statist. Soc. Ser. A. 135, pp. 370-383.

[Pow70a] Powell, M. J. D. (1970a), "A Hybrid Method for Nonlinear Equations", pp. 87-114 of Numerical Methods for Nonlinear Algebraic Equations, edited by P. Rabinowitz, Gordon and Breach, London.

[Pow70b] Powell, M. J. D. (1970b), "A FORTRAN Subroutine for Solving Systems of Nonlinear Algebraic Equations", pp. 115-161 of Numerical Methods for Nonlinear Algebraic Equations, edited by P. Rabinowitz, Gordon and Breach, London.

[Pow70c] Powell, M. J. D. (1970c), "A New Algorithm for Unconstrained Optimization", in Nonlinear Programming, edited by J. B. Rosen, O. L. Mangasarian, K. Ritter, Academic Press, New York.

[Pow70d] Powell, M. J. D. (1970d), "A FORTRAN Subroutine for Unconstrained Minimization, Requiring First Derivatives of the Objective Function", report AERE-R. 6469, A.E.R.E. Harwell, Oxfordshire, England.

[Pow71] Powell, M. J. D. (1971), "Recent Advances in Unconstrained Optimization", Math. Programming 1, pp. 26-57.

[Pow75] Powell, M. J. D. (1975), "Convergence Properties of a Class of Minimization Algorithms", pp. 1-28 of Nonlinear Programming 2, edited by O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York.

[Rei71] Reinsch, C. H. (1971), "Smoothing by Spline Functions. II", Numer. Math. 16, pp. 451-454.

[SaiT78] Saigal, R. and Todd, M. J. (1978), "Efficient Acceleration Techniques for Fixed Point Algorithms", SIAM J. Numer. Anal. 15, pp. 997-1007.

[Sor77] Sorensen, D. C. (1977), "Updating the Symmetric Indefinite Factorization with Applications in a Modified Newton's Method", Report ANL-77-49, Argonne National Laboratory.

[Ste69] Stewart, G. W. (1969), "On the Continuity of the Generalized Inverse", SIAM J. Appl. Math. 17, pp. 33-45.

[Tra64] Traub, J. F. (1964), Iterative Methods for the Solution of Equations, Prentice-Hall, Englewood Cliffs, N.J.

[Wil65] Wilkinson, J. H. (1965), "Error Analysis of Transformations Based on the Use of Matrices of the Form I - $2ww^H$", pp. 77-101 of Error in Digital Computation, Volume 2, edited by L. B. Rall, John Wiley and Sons, New York.

DMG/ck

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 2000

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): On Solving Robust and Generalized Linear Regression Problems.

5. TYPE OF REPORT & PERIOD COVERED: Summary Report -- no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): David M. Gay

8. CONTRACT OR GRANT NUMBER(s): DAAG29-75-C-0024; NSF-MCS78-09525

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Work Unit Number 5 - Mathematical Programming and Operations Research

11. CONTROLLING OFFICE NAME AND ADDRESS: See Item 18 below.

12. REPORT DATE: September 1979

13. NUMBER OF PAGES: 30

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

MRC-TSR-2000

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U.S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D.C. 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Unconstrained optimization, regression, numerical linear algebra, roundoff analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A number of regression problems require finding a parameter vector x^* that minimizes an objective function of the form $\sum_i \rho_i(r_i(x))$, where $r_i(x)$ is the ith component of the (generalized) residual vector $r(x)$ associated with the problem and ρ_i is the ith criterion function. The examples given include (linear and nonlinear) least squares, robust regression, logistic regression, and Poisson regression. These problems have a common structure which may often be worth exploiting, especially in cases where $r(x)$ is a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

221200

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT (Continued)

nonlinear function of x. We study this structure and how to use it. This provides an opportunity to discuss some ideas applicable to general unconstrained optimization and, in particular, to point out the advantages of a model/trust-region approach. For nonlinear $r(x)$, we recommend generalizations of some techniques that have proven worthwhile on nonlinear least-squares problems in which the optimal residual vector $r(x^*)$ may be either large or small. Appendices consider the numerical linear algebra involved in computing a trial change to the parameter vector x and give a new perturbation theorem on linear least squares that is useful in analyzing the numerical behavior of the procedure recommended for computing this trial change.